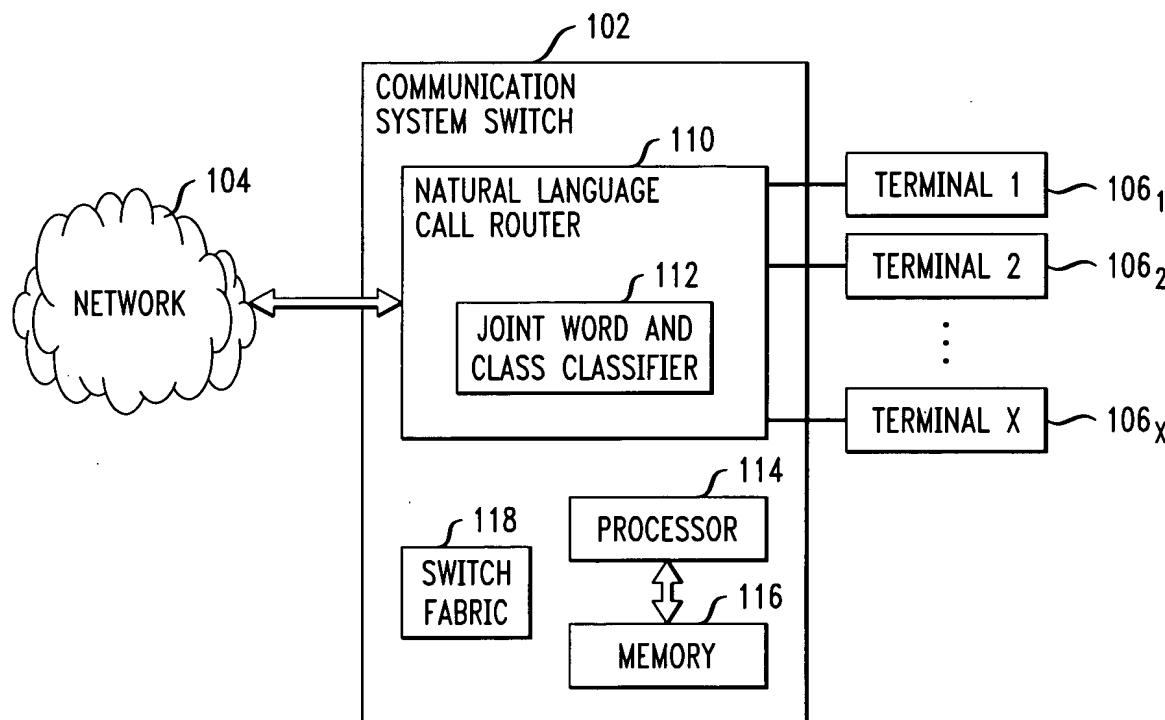


1/5

FIG. 1

100



2/5

FIG. 2

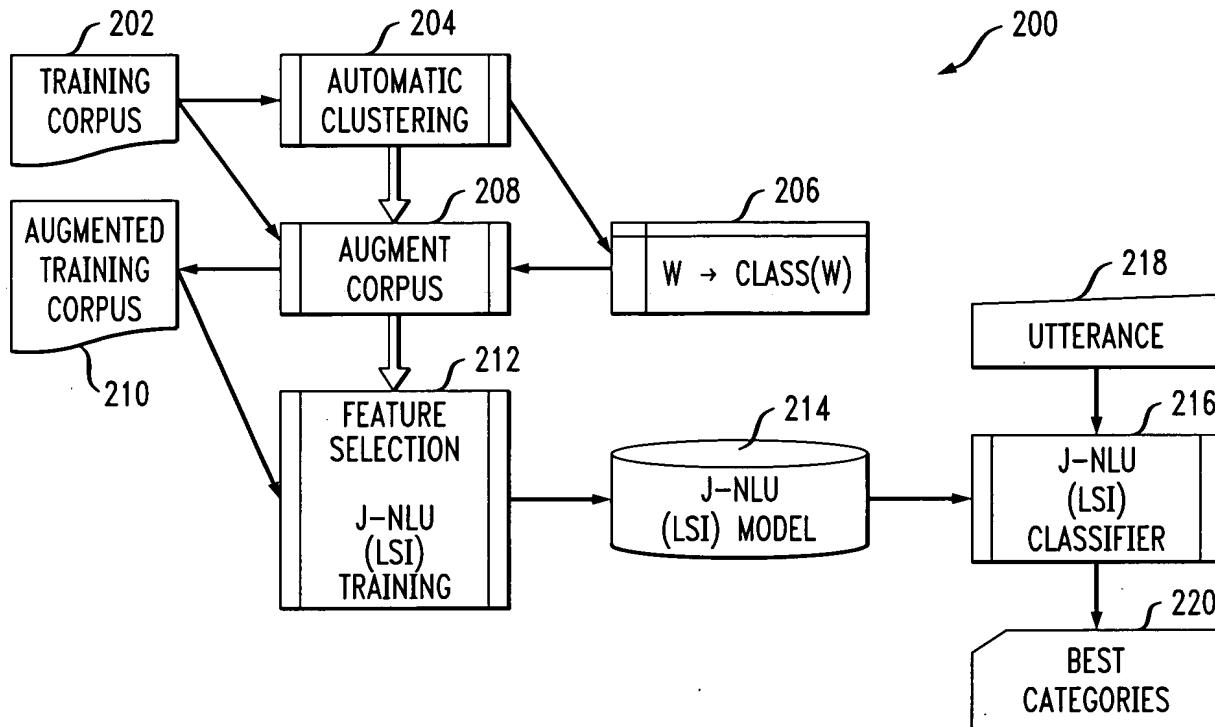
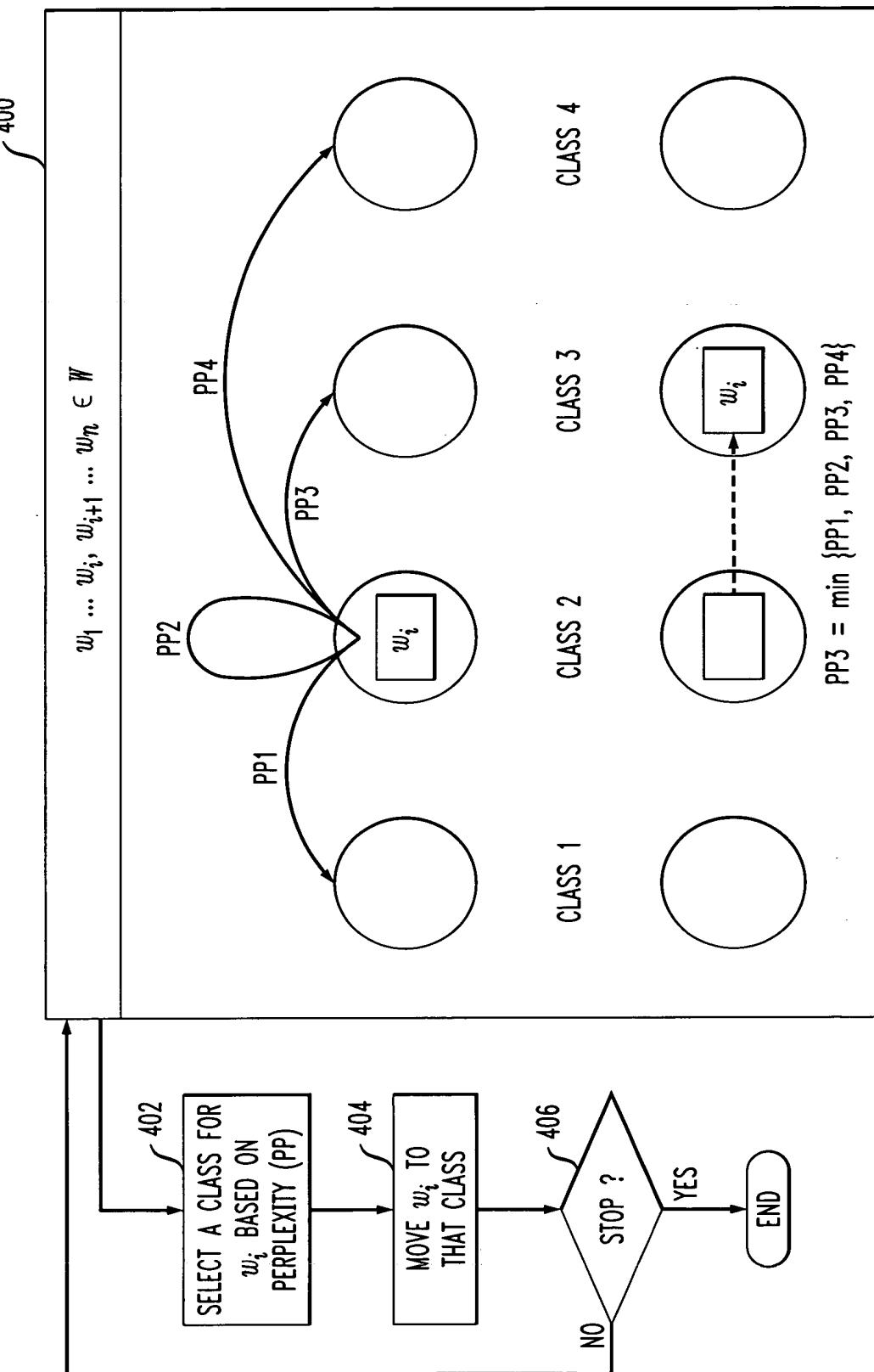


FIG. 3

```
Set up initial word class mapping
Compute the mapping perplexity on the training corpus
Do until some stopping criterion is met
  Do for each word  $w$  in vocabulary  $W$ 
    Remove  $w$  from class  $g_w$ 
    Do for all existing classes  $g$ :
      Compute perplexity as if  $w$  were moved to  $g$ 
    end-do-loop
    Assign  $w$  to the class with the lowest perplexity
  end-do-loop
end-do-loop
Exit
```

3/5

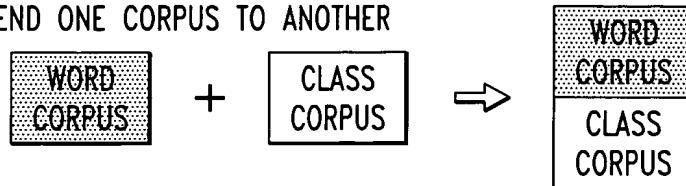
FIG. 4



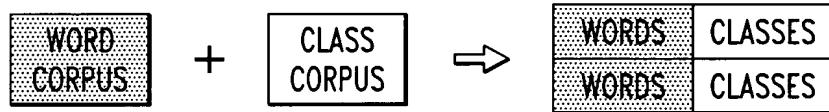
4/5

FIG. 5

APPEND: APPEND ONE CORPUS TO ANOTHER



JOIN: JOIN THE UTTERANCES IN TWO CORPORA



INTERLEAVE: INSERT EACH CLASS AFTER ITS WORD

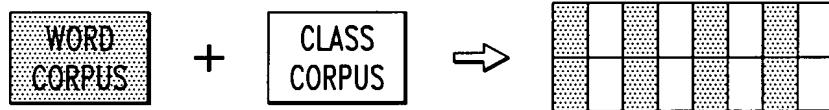


FIG. 6

GIVEN A PARAMETER p :

1. CALCULATE THE IG VALUE OF EACH TERM
2. SORT THE TERMS BY THEIR IG VALUES INTO DESCENDING ORDER
3. SET THE THRESHOLD t TO THE IG VALUE AT TOP p PERCENTILE OF SORTED TERMS
4. SELECT TERMS WITH IG VALUE $\geq t$

5/5

FIG. 7

